

Hölder-type inequalities and their applications to concentration and correlation bounds

Christos Pelekis* Jan Ramon† Yuyi Wang‡

November 24, 2015

Abstract

Let $Y_v, v \in V$, be $[0, 1]$ -valued random variables having a dependency graph $G = (V, E)$. We show that

$$\mathbb{E} \left[\prod_{v \in V} Y_v \right] \leq \prod_{v \in V} \left\{ \mathbb{E} \left[Y_v^{\frac{\chi_b}{b}} \right] \right\}^{\frac{b}{\chi_b}},$$

where χ_b is the b -fold chromatic number of G . This inequality may be seen as a dependency-graph analogue of a generalised Hölder inequality, due to Helmut Finner. Additionally, we provide applications of Hölder-type inequalities to concentration and correlation bounds for sums of weakly dependent random variables.

Keywords: fractional chromatic number; Finner's inequality; Janson's inequality; dependency graph; hypergraphs

1 Introduction

The main purpose of this article is to illustrate that certain Hölder-type inequalities can be employed in order to obtain concentration and correlation bounds for sums of, possibly dependent, real-valued random variables whose dependencies are described in terms of graphs, or hypergraphs. Before being more precise, let us begin with some notation and definitions that will be fixed throughout the text.

A *hypergraph* \mathcal{H} is a pair (V, \mathcal{E}) where V is a finite set and \mathcal{E} is a family of subsets of V . The

*Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001, Belgium, Email: pelekis.chr@gmail.com

†Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001, Belgium, Email: Jan.Ramon@cs.kuleuven.be

‡Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001, Belgium, Email: Yuyi.Wang@cs.kuleuven.be

set V is called the *vertex set* of \mathcal{H} and the set \mathcal{E} is called the *edge set* of \mathcal{H} ; the elements of \mathcal{E} are called *hyperedges* or just edges. The cardinality of the vertex set will be denoted by $|V|$ and the cardinality of the edge set by $|\mathcal{E}|$. A hypergraph is called *k-uniform* if every edge from \mathcal{E} has cardinality k . A 2-uniform hypergraph is a *graph*. The *degree* of a vertex $v \in V$ is defined as the number of edges that contain v . A hypergraph will be called *d-regular* if every vertex has degree d . A subset $V' \subseteq V$ is called *independent* if it does not contain any edge from \mathcal{E} . A *fractional matching* of a hypergraph, $\mathcal{H} = (V, \mathcal{E})$, is a function $\phi : \mathcal{E} \rightarrow [0, 1]$ such that $\sum_{e: v \in e} \phi(e) \leq 1$, holds true for all vertices $v \in V$. The *fractional matching number* of \mathcal{H} , denoted $\nu^*(\mathcal{H})$, is defined as $\max_{\phi} \sum_{e \in \mathcal{E}} \phi(e)$ where the maximum runs over all fractional matchings of \mathcal{H} . The *chromatic number* of a graph G is defined in the following way. A *b-fold coloring* of G is an assignment of sets of size b to the vertices of the graph in such a way that adjacent vertices have disjoint sets. A graph is *(a : b)-colorable* if it has a *b-fold coloring* using a different colors. The least a for which the graph is *(a : b)-colorable* is the *b-fold chromatic number* of the graph, denoted $\chi_b(G)$. The *fractional chromatic number* of a graph G is defined as $\chi^*(G) = \inf_b \frac{\chi_b(G)}{b}$. Here and later, $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$ will denote probability and expectation, respectively.

Let us also recall Hölder's inequality. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let A be a finite set and let $Y_a, a \in A$, be random variables from Ω into \mathbb{R} . Suppose that $w_a, a \in A$ are non-negative weights such that $\sum_{a \in A} w_a \leq 1$ and each Y_a has finite $\frac{1}{w_a}$ -moment, i.e., $\mathbb{E} [Y_a^{1/w_a}] < +\infty$, for all $a \in A$. Hölder's inequality asserts that

$$\mathbb{E} \left[\prod_{a \in A} Y_a \right] \leq \prod_{a \in A} \mathbb{E} [Y_a^{1/w_a}]^{w_a}.$$

This is a classic result (see [2]). In this article we shall be interested in applications of Hölder-type inequalities to concentration and correlation bounds for sums of weakly dependent random variables. We focus on two particular types of dependencies between random variables. The first one is described in terms of a hypergraph.

Definition 1 (hypergraph-correlated random variables). *Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph. Suppose that $\{Y_e\}_{e \in \mathcal{E}}$ is a collection of real-valued random variables, indexed by the edge set of \mathcal{H} , that satisfy the following: there exist independent random variables $\{X_v\}_{v \in V}$ indexed by the vertex set V such that, for every edge $e \in \mathcal{E}$, $Y_e = f_e(X_v; v \in e)$ is a function that depends only on the random variables X_v with $v \in e$. We will refer to the aforementioned random variables $\{Y_e\}_{e \in \mathcal{E}}$ as \mathcal{H} -correlated, or simply as hypergraph-correlated, when there is no confusion about the underlying hypergraph.*

Hypergraph-correlated random variables are encountered in the theory of random graphs (see [8, 10, 16] and references therein). Another type of “dependency structure” that plays

a key role in probabilistic combinatorics and related areas involves the notion of dependency graphs (see [1, 11]).

Definition 2 (Dependency graph). *A dependency graph for the random variables $\{Y_v\}_{v \in V}$, indexed by a finite set V , is any loopless graph, $G = (V, E)$, whose vertex set V is the index set of the random variables and whose edge set is such that if $V' \subseteq V$ and $v_i \in V$ is not incident to any vertex of V' , then Y_{v_i} is mutually independent of the random variables $Y_{v'}$ for which $v' \in V'$. We will refer to random variables $\{Y_v\}_v$ having a dependency graph G as G -dependent or as graph-dependent.*

If $\{Y_e\}_{e \in \mathcal{E}}$ are hypergraph-correlated random variables, then one can define their dependency graph whose vertex set is \mathcal{E} and with edges joining any two sets $e, e' \in \mathcal{E}$ such that $e \cap e' \neq \emptyset$. Hence a set of hypergraph-correlated random variables is graph-dependent. The reader might wonder whether the converse holds true. We will see, using a particular generalisation of Hölder's inequality, that this is not the case (see Example 2.7 below) and so the aforementioned notions of dependencies are not equivalent.

In the present paper we shall be interested in employing Hölder-type inequalities in order to obtain concentration and correlation bounds for sums of hypergraph-correlated random variables as well as for sums of graph-dependent random variables.

The main results are stated in Section 2 and the proofs are contained in Sections 3 and 4.

2 Results

2.1 Hölder-type inequalities

We begin with the following theorem, due to Helmut Finner, that provides a generalisation of Hölder's inequality for hypergraph-correlated random variables.

Theorem 2.1 (Finner [6]). *Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph and let $\{Y_e\}_{e \in \mathcal{E}}$ be \mathcal{H} -correlated random variables. If $\phi : \mathcal{E} \rightarrow [0, 1]$ is a fractional matching of \mathcal{H} then*

$$\mathbb{E} \left[\prod_{e \in \mathcal{E}} Y_e \right] \leq \prod_{e \in \mathcal{E}} \left\{ \mathbb{E} \left[Y_e^{1/\phi(e)} \right] \right\}^{\phi(e)}.$$

Notice that, by applying the previous result to the random variables $Z_e = Y_e^{\phi(e)}$, one concludes $\mathbb{E} \left[\prod_{e \in \mathcal{E}} Y_e^{\phi(e)} \right] \leq \prod_{e \in \mathcal{E}} \{ \mathbb{E} [Y_e] \}^{\phi(e)}$. See [6] for a proof of this result that is based on Fubini's theorem and Hölder's inequality. Alternatively, see [14] for a proof that uses the concavity of the weighted geometric mean and Jensen's inequality. In other words,

the previous result provides a Hölder-type inequality for hypergraph-correlated random variables which is formalised in terms of a fractional matching of the underlying hypergraph. In this article we provide an analogue of Theorem 2.1 for random variables having a dependency graph G . The corresponding Hölder-type inequality is formalised in terms of the b -fold chromatic number of G . More precisely, we obtain the following result.

Theorem 2.2. *Let $\{Y_v\}_v$ be real-valued random variables having a dependency graph $G = (V, E)$. Then, for every b -fold coloring of G using $\chi_b := \chi_b(G)$ colors, we have*

$$\mathbb{E} \left[\prod_{v \in V} Y_v \right] \leq \prod_{v \in V} \left\{ \mathbb{E} \left[Y_v^{\frac{\chi_b}{b}} \right] \right\}^{\frac{b}{\chi_b}}.$$

We prove Theorem 2.2 in Section 3. The proof employs the concavity of the weighted geometric mean and the definition of b -fold chromatic number. In the remaining part of the current section we discuss applications of Theorem 2.1 and Theorem 2.2 to concentration and correlation bounds for sums of hypergraph-correlated random variables as well as for sums of random variables having a dependency graph. We begin with the later case.

2.2 Applications

2.2.1 Dependency graphs

Theorem 2.2, combined with standard techniques based on exponential moments, yields concentration inequalities for sums of random variables having a dependency graph. More precisely, Theorem 2.2 yields a new proof of the following estimate on the probability that the sum of graph-dependent random variables is significantly larger than its mean.

Theorem 2.3 (Janson [10]). *Let $\{Y_v\}_{v \in V}$ be $[0, 1]$ -valued random variables having a dependency graph $G = (V, E)$. Set $q := \frac{1}{|V|} \mathbb{E} [\sum_v Y_v]$. If $t = n(q + \varepsilon)$ for some $\varepsilon > 0$, then*

$$\mathbb{P} \left[\sum_v Y_v \geq t \right] \leq \exp \left(-\frac{2\varepsilon^2 |V|}{\chi^*} \right),$$

where $\chi^* = \chi^*(G_{\mathcal{H}})$ is the fractional chromatic number of G .

See [10], Theorem 2.1, for a proof of this result that is based on breaking up the sum into a particular linear combination of sums of independent random variables. In Section 3 we provide a new proof of Theorem 2.3 which is based on Theorem 2.2. Moreover, under

additional information on the variance of the random variables, we obtain the following Bennett-type inequality.

Theorem 2.4. *Let $\{Y_v\}_{v \in V}$ be random variables having a dependency graph $G = (V, E)$. For every $v \in V$ let $\sigma_v^2 := \text{Var}(Y_v)$ and assume further that $Y_v \leq 1$ and $\mathbb{E}[Y_v] = 0$. Set $S = \sum_v \sigma_v^2$ and fix $t > 0$. Then*

$$\mathbb{P} \left[\sum_v Y_v \geq t \right] \leq \exp \left(-\frac{S}{\chi^*(G)} \psi \left(\frac{t}{S} \right) \right),$$

where $\psi(x) = (1+x) \ln(1+x) - x$.

Let us remark that the previous result is in fact an improvement upon Theorem 2.3 from [10]. Indeed, in [10] Theorem 2.3, the bound $\exp \left(-\frac{S}{\chi^*(G)} \psi \left(\frac{4t}{5S} \right) \right)$ is obtained on the tail probability of Theorem 2.4. Notice that we assume a one-sided bound on each Y_v . The proof of the previous result is similar to the proof of Theorem 2.3; we sketch it in Section 3. In the next section we discuss application of Theorem 2.1 to sums of hypergraph-correlated random variables.

2.2.2 Hypergraph-correlated random variables

In this section we discuss applications of Finner's inequality. We begin by applying Theorem 2.1 to the following question.

Problem 2.5. *Fix a hypergraph $\mathcal{H} = (V, \mathcal{E})$ and let \mathbb{I} be a random subset of V formed by including vertex $v \in V$ in \mathbb{I} with probability $p_v \in (0, 1)$, independently of other vertices. What is an upper bound on the probability that \mathbb{I} is independent?*

Here and later, given a set of parameters in $(0, 1)$, say $\mathbf{p} = \{p_v\}_{v \in V}$, indexed by the vertex set of a hypergraph, we will denote by $\pi(\mathbf{p}, \mathcal{H})$ the probability that \mathbb{I} is independent. The previous problem has attracted the attention of several authors and appears to be related to a variety of topics (see [4, 7, 12, 13, 16] and references therein). A particular line of research is motivated by question about independent sets and subgraph counting in random graphs. In this context, Problem 2.5 has been considered by Janson et al. [12], Krivelevich et al. [13] and Wolfowitz [16]. It is observed in [13] that when \mathcal{H} is k -uniform and d -regular an exponential estimate on $\pi(\mathbf{p}, \mathcal{H})$, can be obtained using the so-called Janson's inequality (see [11], Chapter 2). Additionally, it is shown that under certain "mild additional assumptions" the bound provided by Janson's inequality can be improved to

$$\pi(\mathbf{p}, \mathcal{H}) \leq \exp \left(-\Omega \left(\frac{p|\mathcal{E}|}{(1-p)kd} \right) \right).$$

See [13] and for a precise formulation of the additional assumptions and a proof of this result that is based on a martingale-type concentration inequality. In Section 4 we provide the following upper bound on $\pi(\mathbf{p}, \mathcal{H})$ using Finner's inequality.

Theorem 2.6. *Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph. For each $e \in \mathcal{E}$, let $|e|$ denote its cardinality. Then*

$$\pi(\mathbf{p}, \mathcal{H}) \leq \prod_e \left(1 - \prod_{v \in e} p_v \right)^{\phi(e)},$$

where $\phi : \mathcal{E} \rightarrow [0, 1]$ is a fractional matching of \mathcal{H} . In particular, if the hypergraph \mathcal{H} is k -uniform and $p_v = p$, for all $v \in V$ then

$$\pi(\mathbf{p}, \mathcal{H}) \leq \left(1 - p^k \right)^{\nu^*(\mathcal{H})},$$

where $\nu^*(\mathcal{H})$ is the fractional matching number of \mathcal{H} .

Let us remark that the second statement in Theorem 2.6 has a *monotonicity property*, in the sense that if \mathcal{H}_1 is a superhypergraph of \mathcal{H}_2 then $(1 - p^k)^{\nu^*(\mathcal{H}_1)} \leq (1 - p^k)^{\nu^*(\mathcal{H}_2)}$.

In Section 4 we show that Theorem 2.6 can be seen as an alternative to Janson's inequality. Moreover, using Finner's inequality, one can conclude that the two notions of dependencies given in Definition 1 and Definition 2 are not equivalent.

Example 2.7. *Let G be a cycle-graph on 5 vertices $\{v_1, \dots, v_5\}$ such that v_i is incident to v_{i+1} , for $i \in \{1, 2, 3, 4\}$ and v_5 is incident to v_1 . Let $Y = (Y_1, \dots, Y_5)$ be a vector of Bernoulli 0/1 random variables whose distribution is defined as follows. The vector Y takes the value $(0, 0, 0, 0, 0)$ with probability $\frac{1}{2}(2 - p)(1 - p)^2$, the value $(1, 1, 1, 1, 1)$ with probability $\frac{p^2 + p^3}{2}$, the values*

$$(0, 0, 0, 1, 1), (0, 0, 1, 1, 0), (0, 1, 1, 0, 0), (1, 1, 0, 0, 0), (1, 0, 0, 0, 1)$$

with probability $\frac{p(1-p)^2}{2}$, the values

$$(0, 0, 1, 1, 1), (0, 1, 1, 1, 0), (1, 1, 1, 0, 0), (1, 1, 0, 0, 1), (1, 0, 0, 1, 1)$$

with probability $\frac{p^2 - p^3}{2}$ and the remaining values with probability 0. Elementary, though quite tedious, calculations show that $\mathbb{E}[Y_j] = p$, for $j = 1, \dots, 5$ and that G is a dependency graph for $\{Y_j\}_{j=1}^5$. Now assume that $\{Y_j\}_{j=1}^5$ are \mathcal{H} -correlated, for some hypergraph $\mathcal{H} = (V, \mathcal{E})$. Notice that $|\mathcal{E}| = 5$. If $e_i \in \mathcal{E}$ is the edge corresponding to the random variable Y_i , $i = 1, \dots, 5$, then the fact that $\{Y_i\}_{i=1}^5$ have G as a dependency graph implies that $e_i \cap e_{(i+2) \bmod 5} = \emptyset$, for $i \in \{1, 2, 3, 4, 5\}$. This means that the fractional matching number of \mathcal{H} is at least 2.5 and therefore Theorem 2.1 implies that $\mathbb{P}[Y = (1, 1, 1, 1, 1)] \leq p^{2.5}$. However, the arithmetic geometric means inequality implies $\frac{p^2 + p^3}{2} > p^{2.5}$ and therefore the random variables $\{Y_j\}_{j=1}^5$ are not hypergraph-correlated.

In the same vein as in the previous section, Theorem 2.1 can be employed in order to deduce concentration inequalities for sums of hypergraph-correlated random variables. This has been reported in prior work and so we only provide the statement without proof. In [14] one can find a proof of the following result.

Theorem 2.8 (Ramon et al. [14]). *Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph and assume that $\{Y_e\}_{e \in \mathcal{E}}$ are \mathcal{H} -correlated random variables. Assume further that $Y_e \in [0, 1]$, for all $e \in \mathcal{E}$, and that $\mathbb{E}[Y_e] = p_e$, for some $p_e \in (0, 1)$. Let $\phi : \mathcal{E} \rightarrow [0, 1]$ be a fractional matching of \mathcal{H} and set $\Phi = \sum_e \phi(e)$, $p = \frac{1}{|\mathcal{E}|} \sum_e \mathbb{E}[Y_e]$. If t is a real number from the interval $(\Phi p, \Phi)$ such that $t = \Phi(p + \varepsilon)$, then*

$$\mathbb{P} \left[\sum_{e \in \mathcal{E}} \phi(e) Y_e \geq t \right] \leq \exp(-2\Phi\varepsilon^2).$$

In particular, if d is the maximum degree of \mathcal{H} and $\phi(e) = \frac{1}{d}$, for all $e \in \mathcal{E}$, then the previous result yields the bound

$$\mathbb{P} \left[\sum_e Y_e \geq t \right] \leq \exp \left(-2 \frac{|\mathcal{E}|}{d} \varepsilon^2 \right), \text{ for } t = |\mathcal{E}|(p + \varepsilon).$$

This inequality has also been obtained in Gavinsky et al. [8] using entropy ideas.

3 Proofs - dependency graphs

In this section we prove Theorem 2.2, Theorem 2.3 and Theorem 2.4. The proof of the first theorem will require the concavity of the weighted geometric mean.

Lemma 3.1. *Let $\beta = (\beta_1, \dots, \beta_k)$ be a vector of non-negative real numbers such that $\sum_{i=1}^k \beta_i = 1$. Then the function $g : \mathbb{R}^k \rightarrow \mathbb{R}$ defined by $g(t) = \prod_{i=1}^k t_i^{\beta_i}$ is concave.*

Proof. This is easily verified by showing that the Hessian matrix is positive definite. See [5], or [14] for details. \square

We now ready to prove Theorem 2.2.

Proof of Theorem 2.2. We show that

$$\mathbb{E} \left[\prod_{v \in V} \{Y_v\}^{\frac{b}{x_b}} \right] \leq \prod_{v \in V} \{\mathbb{E}[Y_v]\}^{\frac{b}{x_b}}.$$

The theorem follows by applying this inequality to the random variables $Z_v = Y_v^{\frac{\chi_b}{b}}$. For every color $i = 1, \dots, \chi_b$ let I_i be the set consisting of the vertices that are colored with color i . Note that each I_i is an independent subset of V and every vertex $v \in V$ appears in exactly b independent sets I_i . Therefore,

$$\mathbb{E} \left[\prod_{v \in V} \{Y_v\}^{\frac{b}{\chi_b}} \right] = \mathbb{E} \left[\prod_{i=1}^{\chi_b} \prod_{v \in I_i} \{Y_v\}^{\frac{1}{\chi_b}} \right] = \mathbb{E} \left[\prod_{i=1}^{\chi_b} \left\{ \prod_{v \in I_i} Y_v \right\}^{\frac{1}{\chi_b}} \right].$$

Lemma 3.1 and Jensen's inequality combined with the observation that the random variables $\{Y_v\}_{v \in I_i}$ are mutually independent yield

$$\mathbb{E} \left[\prod_{i=1}^{\chi_b} \left\{ \prod_{v \in I_i} Y_v \right\}^{\frac{1}{\chi_b}} \right] \leq \prod_{i=1}^{\chi_b} \left\{ \mathbb{E} \left[\prod_{v \in I_i} Y_v \right] \right\}^{\frac{1}{\chi_b}} = \prod_{i=1}^{\chi_b} \prod_{v \in I_i} \{\mathbb{E}[Y_v]\}^{\frac{1}{\chi_b}}.$$

Now, using again the fact that each vertex v appears in exactly b sets I_i , we conclude

$$\prod_{i=1}^{\chi_b} \prod_{v \in I_i} \{\mathbb{E}[Y_v]\}^{\frac{1}{\chi_b}} = \prod_{v \in V} \{\mathbb{E}[Y_v]\}^{\frac{b}{\chi_b}}$$

and the result follows. \square

Theorem 2.2 yields a new proof of Theorem 2.3. Let us first recall the following, well-known, result whose proof is included for the sake of completeness.

Lemma 3.2. *Let X be a random variable that takes values on the interval $[0, 1]$. Suppose that $\mathbb{E}[X] = p$, for some $p \in (0, 1)$, and let B be a Bernoulli 0/1 random variable such that $\mathbb{E}[B] = p$. If $f : [0, 1] \rightarrow \mathbb{R}$ is a convex function, then $\mathbb{E}[f(X)] \leq \mathbb{E}[f(B)]$.*

Proof. Given an outcome from the random variable X , define the random variable B_X that takes the values 0 and 1 with probability $1 - X$ and X , respectively. It is easy to see that $\mathbb{E}[B_X] = p$ and so B_X has the same distribution as B . Now Jensen's inequality implies

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\mathbb{E}[B_X|X])] \leq \mathbb{E}[f(B_X|X)] = \mathbb{E}[f(B_X)],$$

as required. \square

We now proceed with the proof of Theorem 2.3.

Proof of Theorem 2.3. Fix $h > 0$ and let $q_v = \mathbb{E}[Y_v]$, for $v \in V$. Using Markov's inequality and Theorem 2.2 we estimate

$$\begin{aligned} \mathbb{P} \left[\sum_{v \in V} Y_v \geq t \right] &\leq e^{-ht} \mathbb{E} \left[e^{h \sum_{v \in V} Y_v} \right] \\ &= e^{-ht} \mathbb{E} \left[\prod_{v \in V} e^{h Y_v} \right] \\ &\leq e^{-ht} \prod_{v \in V} \left\{ \mathbb{E} \left[\exp \left(\frac{\chi_b}{b} h Y_v \right) \right] \right\}^{\frac{b}{\chi_b}} \end{aligned}$$

For $v \in V$ let B_v be a Bernoulli 0/1 random variable of mean q_v . The previous lemma implies

$$\begin{aligned} e^{-ht} \prod_{v \in V} \left\{ \mathbb{E} \left[\exp \left(\frac{\chi_b}{b} h Y_v \right) \right] \right\}^{\frac{b}{\chi_b}} &\leq e^{-ht} \prod_{v \in V} \left\{ \mathbb{E} \left[\exp \left(\frac{\chi_b}{b} h B_v \right) \right] \right\}^{\frac{b}{\chi_b}} \\ &= e^{-ht} \prod_{v \in V} \left\{ (1 - q_v) + q_v e^{\frac{\chi_b}{b} h} \right\}^{\frac{b}{\chi_b}}. \end{aligned}$$

Using the weighted arithmetic-geometric means inequality we conclude

$$\begin{aligned} e^{-ht} \prod_{v \in V} \left\{ (1 - q_v) + q_v e^{\frac{\chi_b}{b} h} \right\}^{\frac{b}{\chi_b}} &\leq e^{-ht} \left\{ \sum_{v \in V} \frac{1}{|V|} \left((1 - q_v) + q_v e^{\frac{\chi_b}{b} h} \right) \right\}^{\frac{b}{\chi_b} |V|} \\ &= e^{-ht} \left\{ 1 - q + q e^{\frac{\chi_b}{b} h} \right\}^{\frac{b}{\chi_b} |V|}. \end{aligned}$$

If we minimise the last expression with respect to $h > 0$ we get that h must satisfy $e^{\frac{\chi_b}{b} h} = \frac{t(1-q)}{q(|V|-t)}$ and therefore, since $t = |V|(q + \varepsilon)$, we conclude

$$\mathbb{P} \left[\sum_{v \in V} Y_v \geq t \right] \leq \left\{ \left(\frac{q}{q + \varepsilon} \right)^{q + \varepsilon} \left(\frac{1 - q}{1 - (q + \varepsilon)} \right)^{1 - (q + \varepsilon)} \right\}^{\frac{b}{\chi_b} |V|} = e^{-\frac{b}{\chi_b} |V| D(q + \varepsilon || q)},$$

where $D(q + \varepsilon || q)$ is the Kullback-Leibler distance between $q + \varepsilon$ and q . Finally, using the standard estimate $D(q + \varepsilon || q) \geq 2\varepsilon^2$, we deduce

$$\mathbb{P} \left[\sum_{v \in V} Y_v \geq t \right] \leq e^{-\frac{b}{\chi_b} |V| 2\varepsilon^2}$$

and the result follows upon minimising the last expression with respect to b . \square

The proof of Theorem 2.4 is similar.

Proof of Theorem 2.4. Fix $h > 0$ to be determined later. As in the proof of Theorem 2.3, Markov's inequality and Theorem 2.2 yield

$$\mathbb{P} \left[\sum_v Y_v \geq t \right] \leq e^{-ht} \prod_{v \in V} \left\{ \mathbb{E} \left[\exp \left(\frac{\chi_b}{b} h Y_v \right) \right] \right\}^{\frac{b}{\chi_b}}.$$

Using an inequality proved in [10] (Inequality (3.7) on page 240), we have

$$\mathbb{E} \left[\exp \left(\frac{\chi_b}{b} h Y_v \right) \right] \leq \exp \left(\sigma_v^2 g \left(\frac{\chi_b}{b} h \right) \right),$$

where $g(a) := e^a - 1 - a$. Summarising, we have shown

$$\mathbb{P} \left[\sum_v Y_v \geq t \right] \leq \exp \left(-ht + \left(\frac{b}{\chi_b} e^{h\chi_b/b} - \frac{b}{\chi_b} - h \right) S \right).$$

Now choose $h = \frac{b}{\chi_b} \cdot \ln \left(1 + \frac{t}{S} \right)$ to deduce

$$\begin{aligned} \mathbb{P} \left[\sum_v Y_v \geq t \right] &\leq \exp \left(\frac{b}{\chi_b} t - S \frac{b}{\chi_b} \left(1 + \frac{t}{S} \right) \ln \left(1 + \frac{t}{S} \right) \right) \\ &= \exp \left(-S \frac{b}{\chi_b} \psi \left(\frac{t}{S} \right) \right). \end{aligned}$$

The result follows upon minimising the last expression with respect to b . □

4 Applications of Finner's inequality

4.1 Proof of Theorem 2.6

In this section we prove Theorem 2.6 and discuss applications of this result to the theory of random graphs.

Proof of Theorem 2.6. Let $X_v, v \in V$, be indicators of the event $v \in \mathbb{I}$. For each $e \in \mathcal{E}$ set $Y_e = \prod_{v \in e} B_v$. Clearly, the random variables $\{Y_e\}_{e \in \mathcal{E}}$ are \mathcal{H} -correlated. Now look at the probability $\mathbb{P}[\sum_e Y_e = 0]$. Notice that if all Y_e are equal to zero, then every edge $e \in \mathcal{E}$ contains a vertex, v , such that $B_v = 0$ and vice versa. This implies that if $\sum_e Y_e = 0$ then

the set of vertices, v , for which $B_v = 1$ is an independent subset of V and vice versa. Therefore

$$\mathbb{P} \left[\sum_{e \in \mathcal{E}} Y_e = 0 \right] = \mathbb{E} \left[\prod_{e \in \mathcal{E}} (1 - Y_e) \right] = \pi(\mathbf{p}, \mathcal{H}).$$

From Theorem 2.1 we deduce

$$\mathbb{E} \left[\prod_{e \in \mathcal{E}} (1 - Y_e) \right] \leq \mathbb{E} \left[\prod_{e \in \mathcal{E}} (1 - Y_e)^{\phi(e)} \right] = \prod_{e \in \mathcal{E}} (\mathbb{E} [1 - Y_e])^{\phi(e)}$$

and the first statement follows. To prove the second statement notice that $\mathbb{E} [Y_e] = p^k$, for all $e \in \mathcal{E}$, and therefore

$$\mathbb{E} \left[\prod_{e \in \mathcal{E}} (1 - Y_e) \right] \leq (1 - p^k)^{\sum_e \phi(e)}.$$

The result follows by maximising the expression on the right hand side over all fractional matchings of \mathcal{H} . \square

4.2 Finner's inequality as an alternative to Janson's

4.2.1 Triangles in random graphs

In this section we discuss comparisons between Finner's and Janson's inequality. Janson's inequality (see Janson [9] and Janson et al. [11, Chapter 2]) is a well known result that provides estimates on the probability that a sum of dependent indicators is equal to zero. It is described in terms of the dependency graph corresponding to the indicators. More precisely, let $\{B_v\}_{v \in V}$ be indicators having a dependency graph G . Set $\mu = \mathbb{E} [\sum_v B_v]$ and $\Delta = \sum_{e=\{u,v\} \in G} \mathbb{E} [B_u B_v]$. Janson's inequality asserts that

$$\mathbb{P} \left[\sum_v B_v = 0 \right] \leq \min \left\{ e^{-\mu + \Delta}, \exp \left(\frac{\Delta}{1 - \max_v \mathbb{E} [B_v]} \right) \prod_v (1 - \mathbb{E} [B_v]) \right\}.$$

Janson's inequality has been proven to be very useful in the study of the Erdős-Rényi random graph model, denoted $\mathcal{G}(n, p)$. Recall that such a model generates a random graph on n labelled vertices by joining pairs of vertices, independently, with probability $p \in (0, 1)$. For $G \in \mathcal{G}(n, p)$ let us denote by T_G the number of triangles in G . A typical application of Janson's inequality provides the estimate

$$\mathbb{P} [G \in \mathcal{G}(n, p) \text{ is triangle-free}] \leq (1 - p^3)^{\binom{n}{3}} \cdot \exp \left(\frac{\Delta}{2(1 - p^3)} \right),$$

where $\Delta = 6\binom{n}{4}p^5$. In this section we juxtapose the previous bound with the bound provided by Finner's inequality.

Proposition 4.1. *Let $G \in \mathcal{G}(n, p)$ be an Erdős-Rényi random graph and denote by T_G the number of triangles in G . Then*

$$\mathbb{P}[T_G = 0] \leq (1 - p^3)^{\frac{1}{n-2}\binom{n}{3}}.$$

Proof. We apply Theorem 2.6. Define a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ as follows. Let $v_i, i = 1, \dots, \binom{n}{2}$, be an enumeration of all (potential) edges in G and consider $v_1, \dots, v_{\binom{n}{2}}$ as the vertex set \mathcal{V} of the hypergraph \mathcal{H} . Let $B_{v_i}, i = 1, \dots, \binom{n}{2}$ be independent Bernoulli $\text{Ber}(p)$ random variables, corresponding to the edges of G , and let $E_i, i = 1, \dots, \binom{n}{3}$, be an enumeration of all triplets of edges in G that form (potential) triangles in G . Define $\mathcal{E} = \{E_1, \dots, E_{\binom{n}{3}}\}$ to be the edge set of \mathcal{H} and let Z_i be the indicator of triangle E_i ; thus $T_G = \sum_i Z_i$. Now form a subset \mathbb{I} of \mathcal{V} by picking each vertex, independently, with probability p . Then the probability that \mathbb{I} is independent equals $\mathbb{P}[T_G = 0]$ and, in order to apply Theorem 2.6, we have to find a fractional matching of \mathcal{H} . Since every vertex of \mathcal{H} belongs to $n - 2$ edges in $\mathcal{E} = \{E_1, \dots, E_{\binom{n}{3}}\}$, we obtain a fractional matching, $\phi(\cdot)$ of \mathcal{H} by setting $\phi(E_i) = \frac{1}{n-2}$, for $i = 1, \dots, \binom{n}{3}$. The result follows. \square

Notice that the bound obtained from Janson's inequality is smaller than the previous bound for values of p that are close to 0, but the previous bound does better for large values of p . Similar estimates can be obtained for the probability that a graph $G \in \mathcal{G}(n, p)$ contains no k -clique, for $k \geq 3$. The details are left to the reader.

4.2.2 Paths of fixed length between two vertices in a random graph

In this section we discuss one more application of Finner's inequality. Let $G \in \mathcal{G}(n, p)$ be a random graph on n labelled vertices. Fix two vertices, say u and v . What is an upper bound on the probability that there is no path of length k between u and v ?

A path of length k is a sequence of edges $\{v_0, v_1\}, \{v_1, v_2\}, \dots, \{v_{k-2}, v_{k-1}\}, \{v_{k-1}, v_k\}$ such that $v_i \neq v_j$. We assume $k \geq 3$, otherwise the problem is easy. Let $\{P_i\}_i$ be an enumeration of all (potential) paths of length k between u and v . Clearly, there are $\binom{n-2}{k-1} \cdot (k-1)!$ such paths. Define the hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ as follows. The vertices of \mathcal{H} correspond to the (potential) edges of G and the edges of \mathcal{H} correspond to the sets of edges in G that form a path of length k between u and v . Hence the probability that there is no path of length k between u and v equals $\pi(p, \mathcal{H})$. In order to apply Theorem 2.6 we have to find a fractional matching of \mathcal{H} and so it is enough to find an upper bound on the maximum

degree of \mathcal{H} . To this end, fix an edge, $e = \{x, y\}$, in G . In case one of the vertices x or y is equal to either u or v , then there are $\binom{n-3}{k-2} \cdot (k-2)!$ paths of length k from u to v that pass through edge e . If none of the vertices x, y is equal to u or v , then we count the paths as follows. We first create a path, P_{k-2} , of length $k-2$ from u to v that does not pass through any of the points x, y and then we place the edge $e = \{x, y\}$ in one of $k-2$ available edges in the path P_{k-2} . Since there are two ways of placing the edge e in each slot of P_{k-2} it follows that the number of paths from u to v that go through edge e is equal to $2(k-2) \cdot \binom{n-4}{k-3} \cdot (k-3)!$. If $k \leq (n-1)/2$ then the later quantity is smaller than $\binom{n-3}{k-2} \cdot (k-2)!$, otherwise it is larger than $\binom{n-3}{k-2} \cdot (k-2)!$. Therefore, if $k \leq (n-1)/2$, the fractional matching number of \mathcal{H} is at least $\frac{\binom{n-2}{k-1} \cdot (k-1)!}{\binom{n-3}{k-2} \cdot (k-2)!} = n-2$. If $k > (n-1)/2$ then the fractional matching number of \mathcal{H} is at least $\frac{(n-2)(n-3)}{2(k-2)}$. We have thus proven the following.

Proposition 4.2. *Let $G \in \mathcal{G}(n, p)$. Fix two vertices u, v in G and a positive integer k . If $k \leq (n-1)/2$ then*

$$\mathbb{P}[\text{there is no path of length } k \text{ between } u \text{ and } v] \leq (1 - p^k)^{n-2}.$$

If $k > (n-1)/2$, then

$$\mathbb{P}[\text{there is no path of length } k \text{ between } u \text{ and } v] \leq (1 - p^k)^{\frac{(n-2)(n-3)}{2(k-2)}}.$$

4.2.3 Degrees

Our paper ends with an estimate on the probability that a $G \in \mathcal{G}(n, p)$ contains no vertex of fixed degree.

Proposition 4.3. *Let $G \in \mathcal{G}(n, p)$ and fix a positive integer $d \in \{0, 1, \dots, n-1\}$. Then the probability that there is no vertex in G whose degree equals d is less than or equal to*

$$\left(1 - \binom{n-1}{d} p^d (1-p)^{n-1-d}\right)^{\frac{n}{2}}.$$

Proof. This is yet another application of Theorem 2.6 so we sketch it. Let v_1, \dots, v_n be an enumeration of the vertices of G . Let the hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ be defined as follows. The vertex set \mathcal{V} corresponds to the (potential) edges of G . The edge set $\mathcal{E} = \{E_1, \dots, E_n\}$ corresponds to the vertices of G . That is, for $i = 1, \dots, n$ the edge E_i contains those $u \in \mathcal{V}$ for which the corresponding edges of G are incident to vertex v_i . The result follows from the fact that $|\mathcal{E}| = n$ and the maximum degree of \mathcal{H} is equal to 2. \square

5 Remarks

As mentioned in Janson [10], there exist collections of weakly dependent random variables that do not have a dependency graph. The dependencies between such collections of random variables can occasionally be described using an independence system. Recall that an *independence system* is a pair $\mathcal{A} = (V, I)$ where V is a finite set and I is a collection of subsets of V (called the *independent sets*) with the following properties (see [3]):

- The empty set is independent, i.e., $\emptyset \in I$. (Alternatively, at least one subset of V is independent, i.e., $I \neq \emptyset$.)
- Every subset of an independent set is independent, i.e., for each $A' \subset A \subset \mathcal{A}$, if $A \in I$ then $A' \in I$. This is sometimes called the *hereditary property*.

Given a set of random variables $\{Y_v\}_{v \in V}$, we say that their joint distribution is *described with an independence system*, say $\mathcal{A} = (V, I)$, if for every $A \in I$ the random variables $\{Y_a\}_{a \in A}$ are mutually independent. Let us remark that this definition includes the case of k -wise independent random variables (see [1], Chapter 16, or [15]). Notice that if $\{Y_v\}_{v \in V}$ are random variables whose joint distribution is described with an independence system $\mathcal{A} = (V, I)$ then $\{v\} \in I$, for all $v \in V$. It is easy to see that if the random variables $\{Y_v\}_{v \in V}$ have a dependency graph then their joint distribution is described with an independence system. However, the converse need not be true. In a similar way as in Section 1, one may define the fractional chromatic number of an independence system as follows.

A b -fold coloring of an independence system $\mathcal{A} = (V, I)$ is a function $\lambda : I \rightarrow \mathbb{Z}_+$ such that $\sum_{A: v \in A} \lambda(A) = b$, for all $v \in V$. The b -fold chromatic number of \mathcal{A} is defined as $\chi_b(\mathcal{A}) := \inf_{\lambda} \sum_{A \in I} \lambda(A)$, where the infimum is over all b -fold colorings, $\lambda(\cdot)$, of $\mathcal{A} = (V, I)$. Finally, the fractional chromatic number of \mathcal{A} is $\chi^*(\mathcal{A}) := \inf_b \frac{\chi_b(\mathcal{A})}{b}$.

With these concepts by hand, one can prove a corresponding Hölder-type inequality using a similar argument as in Theorem 2.2. As a consequence one can obtain tail bounds similar to Theorem 2.3 and Theorem 2.4, the only difference being that the fractional chromatic number of dependency graphs, $\chi^*(G)$, is replaced with the fractional chromatic number of the independence system, $\chi^*(\mathcal{A})$. We leave the details to the reader.

Acknowledgements The authors are supported by ERC Starting Grant 240186 "MiGraNT, Mining Graphs and Networks: a Theory-based approach".

References

- [1] N. Alon, J. Spencer, *The Probabilistic Method*, 3rd Edition, J. Wiley and Sons, New York, (2008).

- [2] V.I. Bogachev, *Measure Theory*, Volume I, Springer, (2007).
- [3] J.A. Bondy, U.S.R. Murty, *Graph Theory*, Graduate Texts in Mathematics 244, Springer, (2008).
- [4] R. Boppana, J. Spencer, *A useful elementary correlation inequality*, Journal of Combinatorial Theory, Series A **50**, p. 305–307, (1982).
- [5] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge UK, (2004).
- [6] H. Finner, *A generalization of Hölder’s inequality and some probability inequalities*, The Annals of Probability, Vol. 20, no. 4, p. 1893–1901, (1992).
- [7] D. Galvin, Y. Zhao, *The number of independent sets in a graph with small maximum degree*, Graphs and Combinatorics **27**, p. 177–186, (2011).
- [8] D. Gavinsky, S. Lovett, M. Saks, S. Srinivasan, *A Tail Bound for Read-k Families of Functions*, Random Structures & Algorithms **47**, p. 99–108, (2015).
- [9] S. Janson, *New versions of Suen’s correlation inequality*, Random Structures & Algorithms **13**, p. 467–483, (1998).
- [10] S. Janson, *Large Deviations for Sums of Partly Dependent Random Variables*, Random Structures & Algorithms **24**, p. 234–248, (2004).
- [11] S. Janson, T. Luczak, A. Ruciński, *Random Graphs*, Wiley, New York, (2000).
- [12] S. Janson, L. Warnke, *The Lower Tail: Poisson Approximation Revisited*, Random Structures & Algorithms, DOI: 10.1002/rsa.20590, (2015).
- [13] M. Krivelevich, B. Sudakov, V.H. Vu, N.C. Wormald, *On the probability of independent sets in random graphs*, Random Structures & Algorithms **22**, p. 1–14, (2003).
- [14] J. Ramon, Y. Wang, Z. Guo, *Learning from networked examples*, Journal of Machine Learning Research, (2015), (to appear), arXiv:1405.2600.
- [15] J.P. Schmidt, A. Siegel, A. Srinivasan, *Chernoff-Hoeffding bounds for applications with limited independence*, SIAM Journal of Discete Mathematics, vol. 8, no. 2, 223–250 (1995).
- [16] G. Wolfowitz, *A concentration result with application to subgraph count*, Random Structures & Algorithms **40**, p. 254–267, (2012).